

**Mad Dog Nativism**  
**Fodor on the Innateness of Concepts**

Fiona Cowie  
Caltech

Final Draft  
September 1997

**NOTE (2015):** This paper was published as 'Mad Dog Nativism,' *British Journal for the Philosophy of Science*, 49 (1998), pp. 227-252.

**Abstract**

Pretty much no-one apart from Fodor believes the radical concept nativism he first articulated in his (1975). Yet that position has stood, virtually unchallenged, for more than twenty years. This neglect is puzzling, for Fodor's arguments against concepts' being learnable from experience remain unanswered, and nativism has historically been taken seriously as a response to empiricism's perceived shortcomings. In this paper, I urge that Fodorean nativism should indeed be rejected, although its deficiencies are not so obvious that they can simply be taken for granted, as they hitherto seemingly have been. After explicating Fodor's position, I examine three arguments, offered by Sterelny and Loar, against it. I explore the possibility that Fodor could avoid these objections by claiming that they misinterpret him as offering a semantic theory of reference, not a psychological theory of concept acquisition. But while I agree that these arguments do apply to causal theories of reference, I show that they apply equally to Fodor's 'brute-causal' account of concept acquisition.

Since Fodor lacks any theoretical machinery to employ in response to them, I conclude that they are fatal to his rabid concept nativism.

---

## **0. Introduction**

Jerry Fodor espouses an especially radical form of nativism about concepts. Describing himself, with uncharacteristic insight, as a 'mad dog nativist' (1984a: 39), Fodor defends the view that "most de facto lexical concepts are innate" (1981: 306).<sup>1</sup> While one might have predicted a flurry of outraged debate in response to such a claim, the philosophical and cognitive science communities have almost completely ignored Fodor's provocation. I find this neglect puzzling.

For Fodor's arguments, developed in his (1975) book *The Language of Thought* and elaborated in his (1981) paper 'The Present Status of the Innateness Controversy', still present a formidable challenge to standard computational accounts of concept-acquisition. Moreover nativism has, at least in the past, always been taken very seriously as a response to perceived failures of empiricist learning theories. Yet his arguments remain unanswered. Although I won't speculate as to why philosophers of mind and psychology, normally so hungry for new controversies, have failed so signally to evince any interest in this tantalizing issue, I propose in this paper to give the mad dog nativist's bone the kind of chewing over that it has long deserved. My aim is to offer some actual arguments in support of doing what most of us apparently did instinctually, namely, rejecting Fodor's rabid nativism out of hand.

Sections 1 and 2 recap Fodor's arguments and elucidate his contention that concepts are innate in the sense that they are 'triggered by' and not 'learned from' experience. Fodor's talk of triggering marks his view that the relation between experience and the concept(s) it occasions is arbitrary, or 'brute-

---

<sup>1</sup> Lexical concepts are expressed by a word, not a phrase (Fodor 1981: 260-1).

causal,' in contrast to the 'rational-causal' links between concepts and experiences that traditional theories of learning rely on. Criticism of this type of nativism — to the extent that it has been taken seriously at all — has centred around its conception of the triggering relation. How, it is asked, could the trigger-concept relation be arbitrary (as Fodor claims) when the relations between concepts and their referents are non-arbitrary? In Section 3, I rehearse Sterelny's (1989) and Loar's (1991) elaborations of this objection. In Section 4, I articulate a distinction between two theoretical tasks, namely, explaining the fixation of reference and explaining concept-acquisition, and in Section 5 I remind the reader of Fodor's counterfactual approach to reference and contrast it with more standard 'historical' theories. Then, in Section 6, I show how a defender of nativism could use a counterfactual semantic theory, and the distinction just mentioned, to counter Loar and Sterelny. None the less, I argue in Sections 7-9, the mad dog nativist's troubles are far from over. Reformulated as objections to his account *qua* theory of concept-acquisition, and not *qua* theory of reference-fixing, analogous difficulties are fatal to Fodor position.

### **1. Fodor's argument for the innateness of most concepts**

In *The Language of Thought* (1975), Fodor argued that computational approaches to psychology presuppose that there is an internal system of representation — a 'language of thought,' or 'Mentalese' — with a structure and richness comparable to those of natural languages. Computation is symbol-manipulation, he points out, so if thought is computational, there must be mental symbols to manipulate. Moreover, he urges, the symbol system used in the mind is importantly language-like: it must possess a combinatorial syntax and semantics, for thought, like language, is both productive (apparently unbounded with regard to its complexity and variety) and systematic (the ability to think one thought brings with it the ability to think other, related thoughts). And finally, he argues, the expressive power of Mentalese must be very rich, for one of the things we use it to do is to learn the vocabulary of our natural language. Psychology sees learning in general as involving the testing of hypotheses, formulated in Mentalese, about the matter being learned. Hence learning a language requires, among other things, that one formulate and test hypotheses about the meanings of its

words. Thus the language of thought must have the resources to represent the meaning of any word in any language: it must be “as powerful as any language one can ever learn” (Fodor 1975: 82).

Fodor's postulation of a Language of Thought is one of the two views about the mind (the other being the Modularity thesis, defended in his (1983)) for which he is best known, and a huge literature has grown up about it. Less remarked, though no less remarkable, is his further claim that the language of thought is not only known, but known innately:

...you can't learn a language unless you already know one..the language of thought is known (e.g. is the medium for the computations underlying cognitive processes) but not learned. That is, it is innate. (1975:65)

We can use Mentalese to learn other things, but, Fodor argues in his (1981), Mentalese itself is unlearnable. Consider the acquisition of its vocabulary, that is, of concepts.<sup>2</sup> In order to learn a concept like, say, BACHELOR, one must formulate and test hypotheses about its content.<sup>3</sup> So in order successfully to learn that concept, one must have available the concepts ADULT, UNMARRIED, and MALE, those being necessary to express the correct semantic hypothesis in this case. Of course, there is nothing especially problematic about this, but consider now an unanalyzable concept, like, for argument's sake, CAUSE. In order to succeed in learning CAUSE, one must again formulate the correct semantic hypothesis. But if CAUSE truly is unanalyzable, then one cannot formulate that hypothesis — except circularly, of course, by using CAUSE itself. But using CAUSE to say what CAUSE means is clearly something one is not in a position to do if what one is doing is *learning* CAUSE! So if learning CAUSE requires us to represent its meaning, and if that is possible, given CAUSE's unanalyzability, just in case we have the concept CAUSE available, then CAUSE couldn't be learned at all. So, while the hypothesis-testing model of acquisition may be applied straightforwardly to what

---

<sup>2</sup> Terminological note: I shall use 'acquisition' and 'attainment' (and cognates) to refer in a theory-neutral way to the process (whether it be learning, triggering, maturation, whatever) whereby concepts become available for general cognitive use.

<sup>3</sup> Another terminological note: following Fodor, I use the upper case when mentioning concepts.

Fodor (1981) calls 'complex' or 'internally structured' concepts like BACHELOR, it is viciously circular if applied to the attainment of 'unanalyzable' or 'unstructured' concepts like CAUSE.

Unstructured concepts, then, aren't learned: they must be acquired in some other way. Now, as Fodor grants, empiricists — his name for proponents of the hypothesis-testing model — have always recognized that some concepts must be unlearned, that is, acquired somehow in virtue of how we happen to be built. But, he alleges, what they did *not* recognize was that what makes a concept unlearnable (hence unlearned) is features intrinsic to the concept itself, namely, its lack of internal semantic structure, its unanalyzability. And here comes the critical step in the argument for nativism: if unanalyzability is what marks off the class of unlearned concepts, then most concepts are unlearned! Years of failed attempts at conceptual analysis indicate that most concepts are like CAUSE and unlike BACHELOR: they firmly resist our efforts to provide them with non-circular definitions.<sup>4</sup>

In sum: if most concepts are unanalyzable, most of them couldn't be learned. Most concepts are unanalyzable. Thus, they must be innate. <sup>5</sup>

## **2. Fodor's acquisition theory**

Although Fodor proclaims that most concepts are innate, he does not mean by this to imply that experience is irrelevant to their acquisition. On the contrary, his view is that experience is "causally necessary and sufficient for the availability of all concepts" (1981: 273). Fodor calls the operation by

---

<sup>4</sup> And even in cases where we do succeed in analyzing a concept, it's arguably implausible to think that people who possess that concept know the definition. As Kripke (1972/1980: 119-21) and Putnam (1975: 226-70) have argued, ignorance and error about the exact boundaries of our concepts are facts of conceptual life. So even if most concepts aren't *unanalyzable*, most concepts, as they are represented in most people's heads, are *unanalyzed*. This is enough to make trouble for the hypothesis-testing model of their acquisition.

<sup>5</sup> In his (1979) Fodor runs a similar argument against empiricist (specifically Piagetian) accounts of the acquisition of the LOT's logical vocabulary. To learn the meaning of a connective, say, or a quantifier, one must learn the truth-conditions of sentences containing it. But while all truth-functional connectives can be defined in terms of 'and' and 'not', the truth conditions of sentences containing quantifiers like 'some' or 'all' cannot be specified in a language that does not already contain them. The innate LOT must therefore be as powerful, not just in terms of its expressive ability, but also as regards its logical resources, as any learnable language.

which concepts are acquired through experience “triggering,” and distinguishes it from learning as follows. Learning a concept is a matter of formulating and testing hypotheses about its content. Our experience in learning bears causally on acquisition to the extent that it confirms or disconfirms those hypotheses. Thus concept-learning is a “rational-causal” 1981: 273) process: the “ambient stimulation” that constitutes the causal input to acquisition “must provide inductive support for the concepts it occasions” (1981: 305). On the Fodorean view, by contrast, acquisition “requires no rational relation between a concept and its causes” (305). Triggering is a non-rational, “brute-causal, i.e. 'merely' empirical” (280) sort of business. In triggering,

...the structure of the sensorium is such that certain inputs trigger the availability of certain concepts. Punkt. Whereas in the case of [learning] there are .. rational-causal processes: the psychological mechanisms which mediate the availability of such concepts realize an inductive logic, and the experiences of the organism stand in a confirmation relation to what is learned. (1981: 273)

Experience, then, causes concepts, but not by 'inductively supporting' semantic hypotheses. Instead, the relation between concepts and their occasioning experiences is arbitrary:

(T)he ethological precedents from other species suggest that such relations may often be extremely arbitrary. There are fledgling ducks for which the operative rule seems to be: if it moves, it's mother. The fact that this experience triggers this concept is surely to be understood on a Darwinian model and not on the model of concept learning. In particular, the observation that the stimulus moves is surely not functioning as data confirmatory of some universal duckish hypothesis about the internal structure of MOTHERHOOD. The point is that the Nativist view allows for any amount of this sort of arbitrariness that you like. (1981: 280)

How, though, is the trigger-concept relation arbitrary? Fodor's references to Darwin and to the ethological concept of triggering suggest that to call that relation 'arbitrary' is not to describe it as random or inexplicable or undetermined. For that movement triggers imprinting and that it's imprinting that movement triggers are none of these things. Movement triggers imprinting because movement was a reliable-enough indicator of motherhood in the ancestral duck environment for ducklings who imprinted on moving things to imprint, mostly, on their mothers.<sup>6</sup> And it's imprinting that movement triggers because ducklings who imprinted on the closest moving things (usually, their mothers) were more apt to survive, reproduce and pass their 'genes for imprinting' to their offspring. Given that duckling behaviors evolved by natural selection, in short, we can understand why imprinting should be under the control of movement. To this extent, then, the movement-imprinting connection is *non*-arbitrary, and Fodor's analogy with the arbitrary triggering of concepts is puzzling.

But there is also a sense in which the movement-imprinting relation *is* arbitrary. Since there are rarely unique solutions to evolutionary problems, properties other than movement could have been recruited as triggers for imprinting. Natural selection will favor more or less any (heritable) mechanism that tends to enhance fitness. So it's just an accident, contingent on what duckish psychologies happened to be around for selection to act upon, that it was movement (not, say, duckbilledness or quacking) that was recruited as imprinting's trigger. Hence although the stimulus-response connection in imprinting looks non-arbitrary now, having been fixed by its evolutionary history, it is none the less highly arbitrary *sub specie aeternitatis*: many properties other than those that were actually recruited as triggers for imprinting could have been selected instead.

In sum, there are two central points on which a Fodorean nativist disagrees with the classical empiricist.<sup>7</sup> First, whereas the empiricist gives an epistemic explanation of why a particular stimulus

---

<sup>6</sup> Those in a Lorenz-infested environment were, as is well known, SOL.

<sup>7</sup> This account addresses the theory Fodor outlines in his (1975) and (1981). Given his more recent rejections of attempts to apply evolutionary theory to questions in the cognitive sciences ("It is, I think, unlikely...that Darwin is going to pull Brentano's chestnuts out of the fire" (1990: 70); cf. also (1998, Ch. 6)), it's unclear how happy he would be with this reading. In my (forthcoming), Ch. 4, I

is the cause of a certain concept (redness is the cause of RED because reddish experiences are evidence of what RED means), the Fodorean holds that there is no such explanation: the cause-concept relation is brute-causal, not rational-causal. Emus don't cause EMU because experiences of emus (dis)confirm our hypotheses about what that concept means. Instead, it's just an evolutionary accident, an artifact of the way we were built, that those things trigger that concept.

The nativist and the empiricist disagree, secondly, on the question of whether or not our experiences are sufficient to occasion acquisition of concepts. Fodor thinks that, given how we're built, experience alone is sufficient: "triggering of the sensorium is, normally, causally necessary and sufficient for the availability of all concepts." (1981: 273) The empiricist, by contrast, thinks that experience by itself is not sufficient. Something else is causally implicated in our acquisition of that concept, namely, our formulation and test of hypotheses about its meaning. To put the contrast bluntly, but not, I think, inaccurately: Fodor's nativist thinks that Mother Nature and the world do all the causal work in concept-acquisition, whereas the empiricist thinks that we have to do some work as well.<sup>8</sup>

### **3. Three objections**

---

defend this interpretation of his position, and in Ch. 5 I show how it relates to his current views, including in particular his apparent repudiation of nativism, not addressed in this paper, in his (1998).

<sup>8</sup> The question arises how Fodor's nativism looks in the light of connectionist approaches to concept-acquisition. Can connectionism provide an empiricist alternative to the learning theory denigrated by Fodor, one that is immune to his objections? Or do connectionist models merely implement (rather than competing with) the accounts Fodor describes? How a Fodorean should answer these questions depends on the details of the connectionist account, for connectionism per se is neutral with respect to the nativist-empiricist dispute. If the connectionist were to explain acquisition of PLATYPUS, say, by describing a net in which the initial weights were set such that platypus-ish inputs would result in a stable and characteristic pattern of firing, then Fodor could properly claim that the connectionist's model simply implements the triggering (nativist) theory he proposes. If, on the other hand, the connectionist took a network with initially randomly set weights and established a PLATYPUS-pattern through use of some kind of learning algorithm (e.g. back-propagation), then, I suspect, Fodor would claim that the network implemented the hypothesis-testing model of the empiricist (and was flawed for the reasons that he gives). Whether or not connectionist networks in fact display the kind of semantic constituency that would allow Fodor to make this latter argument remains a hotly-debated topic (cf. Fodor (1998), Fodor and Pylyshyn (1988), Smolensky (1988)) and I see little prospect that the issue of how connectionist and classical conceptions of computation are related will be resolved any time soon.



In this section I survey the three most substantive objections to Fodor's innateness about concepts that I have found in the literature. In Sections 4 and 5, I argue that they are all unsuccessful.

Sterelny (1989: 126f.) denies that the triggers of concepts are arbitrary even in the ways that the causes of imprinting are. He notes that while any moving object will suffice to occasion imprinting, it's untrue that any old thing will suffice to make PLATYPUS available. Only platypuses will do. For, Sterelny argues, the reference of a concept is fixed by what triggers it. So, if what triggers PLATYPUS is not a platypus, then the concept acquired could not be the concept *of* a platypus. Thus, whereas the duckling can imprint in response to a non-motherish trigger, like Konrad Lorenz, the acquisition of PLATYPUS must be caused by a platypus. Hence the stimulus-class for concepts is vastly more restricted than that for imprinting and it is false that nativism could 'allow for any amount of arbitrariness that you like' in the cause-concept relation.

Sterelny also confronts Fodor with a second problem, posed originally in Sterelny (1983) as a difficulty for what Devitt and Sterelny (1987) call 'pure-causal' theories of reference. He calls it 'The *Qua* Problem.' According to a pure-causal theory, the reference of a concept is fixed by the causal connections between it and its referent. In the paradigm (that is, ostensive) case, EMU refers to emus because those things cause tokenings of that concept.<sup>9</sup> Sterelny points out that any concept-tokening caused by an emu is also caused by a bird, a mammal, a vertebrate, an animal, and so on. So why does the concept that emus cause refer *to emus*, rather than to birds, or mammals, or vertebrates, etc.? If EMU refers to emus because its tokenings are caused by emus, there must be something that determines that it is *qua emu* that those animals are causing tokenings of that concept. The *Qua* Problem is the problem of specifying what that 'something' is.

In his (1989), Sterelny applies the *Qua* problem to Fodor's nativism. What makes it the case, he asks, that the concept acquired in response to an emu-trigger is the concept of an emu, and not the concept

---

<sup>9</sup> Cf. Devitt (1981). Pure causal theories contrast with, for example, description theories, according to which the reference of a concept is determined by an associated description.

of any of the other kinds of which emus are members? Sterelny suggests that it's a prototype or stereotype that we associate with the concept that determines which of its many causes fixes its reference. Roughly, it's because we're thinking of a large, flightless bird native to Australia that emus cause EMU and not, say, VERTEBRATE. Thus, he contends, the concept-acquisition mechanism could not be 'brute-causal': 'rational-causal' factors, that is, additional psychological states and/or processing, are implicated in acquisition too.

A final objection to Fodor's nativism, raised by both Sterelny (1989) and Loar (1991: 122ff.), emphasizes that many concepts are not acquired ostensively. While some people, perhaps, acquire PLATYPUS in confrontation with a platypus, most people have never had a platypus-experience and acquired PLATYPUS instead from hearing talk about platypuses, seeing pictures of platypuses, reading National Geographic, browsing the World Wide Web, or whatever. Devitt (1981) calls this 'reference-borrowing': in case we are not directly acquainted with the referent of a concept, we can none the less acquire a concept with that reference by being causally connected (in the right way) with someone who is so acquainted. Loar calls concepts acquired in this manner 'deferential' concepts: we "implicitly take (their) reference to be determined by the language (we) speak" (1991: 120).<sup>10</sup> Whatever you call it, non-ostensive acquisition is a problem for Fodor. How could a brute-causal mechanism, such as he postulates, 'know' to output the concept PLATYPUS and not, for example, THE SOUND /PLATYPUS/ or WORD MEANING PLATYPUS, in response to a trigger comprising someone's uttering 'platypus'? Both Sterelny and Loar argue that further mental states must be involved in selecting the first output over the others. Sterelny surmises that it is an intention to borrow reference that determines that, rather than picking the 'obvious' causal chain linking the acquired concept with its trigger (viz., the sound /platypus/), the acquisition mechanism selects as the determinant of reference the more complex and socially mediated causal chain connecting that concept with platypuses. For Loar, it is a 'guiding conception', or, roughly, an intention that the

---

<sup>10</sup> Loar's conception of deferential concepts is narrower than Devitt's (and Sterelny's, incidentally), for whereas the notion of reference-borrowing applies in cases where the occasioning experience is neither ostensive nor linguistic (as when I see a picture of a platypus), Loar's main concern is with concepts acquired through linguistic experience.

concept refer to whatever the English word 'platypus' refers to, that does the trick. While neither of these proposals has been fleshed out, they share their insistence that non-ostensive acquisition must be psychologically mediated. As a corollary, then, they imply that acquiring 'deferential' or 'borrowed' concepts is not a brute-causal matter. It is eminently 'rational-causal,' or psychological, in nature. Lexical or not, concepts acquired via reference-borrowing could not be innate in Fodor's sense.

#### **4. Theories of reference vs. theories of concept-acquisition**

These objections to Fodor's nativism suffer from a polemical defect, however, in that they tacitly rely on a picture of reference that Fodor now rejects. They are articulated against what I'll call an 'historical' conception of reference, one that assumes that the reference of a concept is fixed by its actual relations (causal, informational, etc.) with things in the world. As a consequence, they suppose that if what is acquired as a result of a triggering experience is the concept F-NESS, then it must have been an F that triggered the acquisition mechanism. For on an historicist view, if a concept were triggered by something that is not F, that concept could not refer to Fs, in which case, it would not be the concept F-NESS.

Thus, Sterelny argues, the triggers for concepts aren't arbitrary, because given an historical semantics, concepts' reference (and hence their identity) depends on the actual nature of the triggering stimulus. Thus also arises the *Qua* Problem. If a given triggering experience is going to cause acquisition of EMU, then it must be *qua* emu that the experience is doing the triggering. For if it's not *qua* emu that an emu triggers the acquisition device, then (if one's semantics is historical) the concept acquired wouldn't refer to emus. And so, finally, do worries about reference-borrowing originate. On an historical conception, concepts refer to what *actually* causes them, that is, on Fodor's view, to their triggers. But then how is reference-borrowing — a process in which concepts come to refer not to their triggers (words, pictures, etc.) but to what their triggers *represent* — possible at all? If concepts' reference is fixed by what causes them, in sum, then there must be more

to the causal story than Fodor is telling us. Concept-acquisition could not be the kind of brute-causal, stimulus-driven process that he describes.

In response to such objections, Fodor might insist on there being a distinction between theorizing about reference and theorizing about concept-acquisition. Theories of reference, or of representation more generally, seek to explain in virtue of what a particular object (word, concept, picture, or other symbol) means what it does. What makes the English word 'cat' refer to cats? In virtue of what does the concept PLATYPUS represent platypuses? Why is this photo a picture of the Notre Dame? Theories of reference thus seek to answer a metaphysical question: what makes it the case that a particular thing is *about* something else?

Theories of concept-acquisition address a quite different question. Rather than asking: In virtue of what does the concept PLATYPUS refer to platypuses? theories of concept-acquisition ask: How do we come to have the concept PLATYPUS, a representation referring to platypuses, in our heads? This is an aetiological question about (ostensibly) psychological processes: what are the processes responsible for our heads' coming to contain objects with the sorts of representational properties that it is the job of theories of reference to explain?<sup>11</sup>

Furthermore, the nativist might continue, the fact that explaining reference and explaining acquisition are in principle distinct theoretical tasks constitutes a loophole through which he can slip to evade Sterelny's and Loar's objections. For although *historical* semantic theories entail that the conditions under which concepts are acquired are also those that fix reference, there are other theories of reference, including the counterfactual theory proposed by Fodor, that do not have this consequence. Hence, as we will see in Sections 5 and 6, Fodor is at liberty to agree with Sterelny and Loar that reference-fixing is not a brute-causal matter, while maintaining nonetheless that concept-acquisition is.

---

<sup>11</sup> Compare: "What justifies your claim that the man disporting himself in this photo is my Uncle Bob?" vs. "How did a photo of Uncle Bob *in flagrante* come to be in your possession?"

## **5. Historical vs. counterfactual contents**

Historical semantic theories imply that reference is fixed by whatever causes acquisition of a concept. Pure-causal theories of the kind propounded by Kripke (1972/1980), Putnam (1975), Devitt (1981), and (with modifications) Devitt and Sterelny (1987), for example, see the reference of a term (or concept) as being fixed by the causal connections holding between it and the object in the presence of which it is introduced. Thus, if triggers cause the first tokenings of concepts, concepts must refer to their triggers. Covariational causal (or 'causal-informational') theories, of the sort defended by Dretske (1981) and explored also at one time by Fodor (1984b, 1990b), hold that the reference of a term is a matter of what it carries information about, where a term carries information about whatever its tokening is *de facto* causally correlated with. Since concepts covary with their triggers, they must carry information about, and therefore refer to, the triggering property.

Fodor, however, came to reject historical semantic theories such as these on the grounds that they fall prey to what he calls 'The Disjunction Problem.'<sup>12</sup> Suppose, the argument goes, that one night you glimpse a skunk slipping through the garden, and think 'There's the cat!', i.e., you token CAT!! We want to say that your thought was mistaken: you thought it was the cat, when really, it was a skunk. Given an historical account of what CAT means, however, we can't say this. For if concepts refer to what in fact causes them (or to what they in fact covary with), then, if your concept was caused by (/covaries with) a skunk, it must refer to skunks. So, assuming that tokens of CAT are also caused at times by cats (as when it really is Bruce, out and about doing his cat stuff), what CAT refers to is cats-or-skunks. So you did not think mistakenly that the cat was in the garden. Instead, you thought, correctly, that a cat-or-skunk was there. Perceptual error, it would seem, is impossible.

---

<sup>12</sup> See, e.g., Fodor (1987) and (1990a: 51-87).

Fodor holds, not unreasonably, that if a theory of reference entails omniscience, then it must be false.<sup>13</sup> So, he urges, instead of focusing on concepts' actual causal (/informational etc.) relations with the world in developing our semantic theories, we should focus instead on their counterfactual relations. If we could somehow make the counterfactuals count, then what would be important to the reference of CAT is not that concept's *actual* relations — be they with cats, skunks, or cats-or-skunks — but rather the relations it *would* have with cats in more ideal semantic circumstances. If CAT's reference were fixed by what tokenings of CAT would (under ideal circumstances) be caused by, viz., cats, then it wouldn't matter one whit that in some circumstances CAT tokens are caused by non-cats. So long as the concept satisfies the relevant counterfactuals, its reference to cats is secure, and the disjunction problem is defused.

The counterfactuals that, on Fodor's view, fix conceptual contents are sustained by a network of natural laws connecting tokenings of the concept with the property to which the concept refers.

Schematically:

A representation R expresses the property P in virtue of its being a law that things that are P cause tokenings of R (in, say, some still-to-be-specified circumstances C). (Fodor 1998: 12, italics removed.)

That is, on his view, CAT refers to cats because there exists a nomic relation between the property being a cat and the property being a cause of CAT-tokens, and *not* because of any actual, historical connections between that concept and instantiations of cat-hood in the world. Of course, this basic picture needs elaboration if it is to deal with the disjunction problem (and, as we'll see, Loar and Sterelny). For if skunks can cause CAT-tokens (as in the story above), there must be a law that permits this too. That is, there must be laws connecting CATs with being a skunk in addition to the

---

<sup>13</sup> The disjunction problem works best against the simplest versions of historical theories. Unsurprisingly, proposals abound in the literature as to how to deal with misrepresentation. I shall not deal with such them here.

laws connecting CATs with being a cat. In which case, according to the schema above, CAT refers to skunks as well as cats, and the disjunction problem remains unsolved.

Elaborating the basic counterfactual story, however, has proven somewhat tricky — so tricky, indeed, that Fodor tends now to leave the semantic details more or less to their own devices.<sup>14</sup> But we can get an inkling of what kind of explication is needed by looking at one of his older formulations of the counterfactual theory. In his (1990a, esp. 93ff.), Fodor postulated an asymmetry among the laws governing tokenings of CATs. He maintained that whereas the ability of cats to cause CATs is independent of any ability that skunks might have to cause CATs, the converse is not the case: skunks cause tokenings of CAT only because cats do. This suggests that, if there are nomic relations between other properties and tokenings of CAT, then those other relations 'asymmetrically depend' on the cat-CAT relation: the skunk-CAT law exists in virtue of there being a cat-CAT law, but not *vice versa*. This asymmetry, Fodor argues, is semantically significant. For reference depends only on the counterfactuals sustained by the primary (cat-CAT) law; secondary laws (such as the skunk-CAT law) are beside the semantical point. Thus, he argued, you did make a mistake when you saw the skunk and thought CAT! Your CAT-token referred to cats, and not to skunks-or-cats, because skunks only cause tokenings of CAT because cats do, that is, because the cat-CAT law is, as I'll say, 'semantically primary'.

I myself have serious doubts about asymmetric dependence and, judging by the conspicuous absence of asymmetry-talk in his recent writings, so has Fodor.<sup>15</sup> But we don't need to pursue those doubts here. What was important about the asymmetric dependence story was that it offered Fodor a way

---

<sup>14</sup> Cf. the semantic 'theory' propounded in his (1998: 12): "Meaning is information (more or less)...what bestows content on mental representations is something about their causal-cum-nomological relations to things that fall under them: for example, what bestows upon a mental representation with the content *dog* is something about its tokenings being caused by dogs. Just how this works depends, of course, on what sort of causal-cum-nomological covariation content is and what sorts of things you think that concepts represent."

<sup>15</sup> It is (putting it mildly) unclear what it is for one law to be 'dependent' on another; and it's unclear too why we should take Fodor's word for how the putative dependencies go. The passage quoted in n. 14 above reflects Fodor's new-found ambivalence about the asymmetric dependence view.

of picking out some of the nomological relations enjoyed by a concept as being semantically primary. It offered, that is to say, a way of distinguishing those laws which supply the counterfactuals that are relevant to the fixation of reference from the 'semantically secondary' laws — those that are irrelevant to reference — and this is what gets him around the disjunction problem. Although there are presumably lots of laws subsuming the myriad causal relations of a concept like CAT, only one of those laws is relevant to what CAT means.

But more significantly for our purposes, and as we'll see in the next section, Fodor's adoption of a counterfactual semantics offers also to save his nativist theory of concept acquisition from the depredations of Sterelny and Loar.

## **6. Answering Sterelny and Loar**

We saw in Section 5 how adopting a counterfactual semantics (and finding some way of distinguishing between semantically primary and semantically secondary laws) dissolves the disjunction problem. This happy result reflects the more general capacity that a counterfactual semantics has to divorce a concept's reference from its aetiology. On this view it's not what *does* happen, but what *would* happen, that is important to fixing reference. In particular, and most importantly for our purposes, it's not necessarily *what actually happens during concept acquisition* that fixes concepts' reference: given a counterfactual semantics, the causal factors implicated in a concept's arising in the mind need have nothing to do with how its reference is determined. For the causal process under which a concept arises may be covered by a semantically secondary law, in which case that causal process is irrelevant to the concept's identity. Reference is secured by the semantically primary law, and that the acquisition process happens to be covered by a secondary law is thus of no semantical interest.

So, first, it doesn't matter to the identity of the concept attained whether triggers are arbitrary or not. So long as the law connecting platypuses with PLATYPUSES is primary, anything at all could cause tokenings of that concept. A fortiori, anything could cause the *first* tokenings of that concept. Fodor



(1981) was right after all: given a counterfactual semantics, the triggering stimuli for concepts can in principle be “as arbitrary as you like” (1981: 280) — even if, as Sterelny correctly observes, in practice they are not.<sup>16</sup>

Nor, secondly, does it matter *qua* what a trigger causes the acquisition of a concept. Sterelny claimed that it was important that it should be *qua* platypuses that platypuses trigger PLATYPUS, for otherwise, how could that concept succeed in referring to those animals? On the counterfactual view, though, PLATYPUS refers to platypuses in virtue of the platypus-PLATYPUS law: the laws connecting monotremes, vertebrates and so on with PLATYPUSes are semantically secondary. So platypuses could trigger PLATYPUS *qua* the things that wowed the Royal Society in 1779 and, so long as the platypus-PLATYPUS law is primary, that concept's reference to platypuses is secure.

And nor, finally, does deference or reference-borrowing constitute a problem. So long as the ability of /platypus/ or TV shows or National Geographic articles to cause tokenings of PLATYPUS is secondary to the ability of platypuses to do so, any of these things could trigger that concept. What concept is acquired depends not on what the trigger is, but rather on what the primary laws governing tokenings of that concept happen to be.

For a counterfactual semanticist such as Fodor, in other words, explaining reference is one thing and explaining acquisition is another. Hence he may grant (with Loar and Sterelny) that the causal transactions underpinning concept-acquisition do not suffice to determine concepts' reference, while resisting their further inference that therefore, those transactions could not suffice for acquisition. The process of concept-acquisition may be brute-causal, even though that by which reference gets determined is not. Hence Fodor can evade the Sterelny-Loar objections to his acquisition theory.

---

<sup>16</sup> This issue – that in fact they're not – is taken up in Fodor (1998); see also Cowie (forthcoming), chapter 5.

Which is not to say that those objections are uninteresting. On the contrary, the points made by Loar and Sterelny have a significance that goes far beyond the assessment of Fodor's particular position. For what their arguments show is that nativism about concept-acquisition is incompatible with a historicist semantic theory. If one takes the view (as many do) that the conditions under which concepts are acquired are also those under which their reference is fixed, then one cannot also hold that acquiring concepts is a brute-causal matter. Sterelny and Loar thus rule out completely one way of being a nativist about concepts. What I have argued in this section, though, is that there remains another option for the nativist to explore. By eschewing an historical semantics, one can deny the assumption, upon which the Sterelny-Loar arguments turned, that 'meaning is method of acquisition'. One can claim that the reference of a concept is determined by the relations (actual or possible) between concepts and objects that are described in the primary laws. Which means that a nativist may agree with Sterelny (and Loar) that the brute-causal processes he describes do not determine reference, yet maintain that this in no way impugns his claim that they are the processes by which concepts are acquired.

In the next sections, we will grant for the sake of argument that reference is taken care of in the way Fodor suggests.<sup>17</sup> The question still remains: is the acquisition mechanism 'brute-causal' rather than 'rational-causal'? I will argue that despite his success in answering Sterelny and Loar, Fodor's views on concept-acquisition are still in serious trouble.<sup>18</sup>

---

<sup>17</sup> This, of course, is a largish assumption. Finding a way to distinguish semantically primary from secondary laws is a serious problem for the counterfactual semanticist. For the primary laws must be primary in virtue of some feature they have *qua* laws (such as whether or not they are dependent on other laws, as in the failed asymmetric-dependence account). Picking out the primary laws *qua* the laws that, were they primary, would secure CAT's reference to cats, begs the question. If your metaphysics is going to do your semantic work for you, there's got to be some reason to believe that the metaphysical facts are as you say they are other than that the metaphysical facts' being that way would get your semantical dirty work done.

<sup>18</sup> The arguments of this section were inspired by certain of Fodor's remarks in his (1990a) and (1991). In his (1990a), he lauds Skinner's semantics because for Skinner, all that matters are relations of "[nomic covariance] between symbols and their denotations. It doesn't matter how that covariance is mediated; it doesn't matter what mechanisms sustain the covariation" (1990a: 56), and argues that just because the covariance is mediated, this doesn't mean that the mediation helps to determine the meaning. In his (1991) response to Loar, Fodor writes: "Ostensive definitions, 'guiding conceptions' and the like may be among the mechanisms that occasion or sustain such

## **7. The issues recast**

Before proceeding, we should take a moment to clarify the task faced by a Fodorean nativist in giving an account, not now of reference, but of concept- acquisition. His contention, recall, is that acquisition is a 'brute-causal' rather than a 'rational-causal' process. Given how we have been 'wired up', experience is both necessary and sufficient for the availability of all concepts. The necessity of experience implies that you don't get an object like ECHIDNA (that is, a symbol referring to echidnas) in your head in the absence of any experience. Its sufficiency implies that experience is all you need for the acquisition of ECHIDNA. No additional 'top-down' or psychological processing (such as the formulation and test of semantic hypotheses) is implicated in that concept's attainment.

In explicating the 'brute-causality' of the acquisition mechanism, Fodor claimed that the experiences occasioning acquisition of a concept 'can be as arbitrary as you like'. While this may be true in principle (that is, an asymmetric-dependence theory of reference allows for indefinitely much arbitrariness in the class of triggering stimuli, as I argued in Section 4), the experiences that as a matter of fact occasion acquisition of a concept like ECHIDNA are of two broad types. We do not acquire ECHIDNA as a result of sitting in a lawn chair, or eating a curry, or watching a Dodgers game. Instead, that concept is overwhelmingly usually acquired either through ostension, or (perhaps the more common way) through reference-borrowing: through hearing someone talk about echidnas, watching TV shows about echidnas, reading books about echidnas and so on. Let us look in a little more detail at each of these processes, as described in Fodorean terms.

In the case of reference-borrowing, let's suppose, I hear the sound /echidna/. This on the Fodorean view should suffice for my acquiring the concept ECHIDNA. That is, assuming that we would not want to deny that concepts are got through reference-borrowing (and denying this would be implausible since it would entail that most of us don't have most of the concepts that, in fact, we do),

---

onomic relations [i.e. those definitive of reference]; but they aren't constitutive of semanticity; only the nomic relations themselves are." (p. 287).

hearing someone talk about echidnas should suffice for my acquiring the ability to token the symbol ECHIDNA such that (i) there is a nomological relation between being an echidna and being a cause of tokenings of ECHIDNA and (ii) the laws in virtue of which other properties (such as the property being the sound /echidna/ ) cause tokenings of ECHIDNA depend asymmetrically on the existence of the primary law specified in (i). In the case of ostension, let's say, I see an echidna. This again, according to Fodor, should suffice for my acquiring the ability to token a concept, ECHIDNA, in virtue of the primary law specified in (i) above.

One can thus view the difference between ostension and reference-borrowing as boiling down to a difference in which law the acquisition process is subsumed under. In the ostensive case, it is the primary law that is operative in that process; in the case of reference- borrowing, it is a secondary law that is at work. (In both cases, though, it's still the primary law that secures reference. Remember that we are explaining the aetiology of the concept, not the fixation of its reference.)

The nativist's task, then, is to describe for us a device embodying brute-causal processes which can be covered, variously, by either of the primary and secondary laws just mentioned. That is, while it's perfectly acceptable, when dealing with reference per se, to abstract from the mechanisms that sustain the networks of nomological relations Fodor envisages, such abstraction is emphatically not acceptable when concept-acquisition is at issue. For it is precisely an account of those mechanisms that an acquisition theory must provide.

## **8. The mechanism of ostensive acquisition**

First, let's take ostension. In this case, we must conceive a brute-causal mechanism such that seeing (or otherwise coming into sensory contact with) an echidna suffices to cause it to deliver ECHIDNA to the mind. No problem, you think: we assumed above that in ostensive acquisition, the acquisition process was subsumed under the primary law. So let's just stipulate that it is *qua* echidnas that echidnas cause ostensive acquisition of ECHIDNA: it is the property being an echidna that is causally

implicated in the mechanism's delivery of that concept. There is nothing inherently odd in supposing that, just as a Coke machine is designed such that the property being coins of a certain size and weight causes delivery of a Coke, we were 'designed' — that is, we evolved — such that being an echidna causes delivery of ECHIDNA.

However, as Sterelny pointed out in his presentation of the *Qua* problem, anything that has the property being an echidna also has the property being a vertebrate. Which raises the question: is the acquisition mechanism also sensitive to that property? Here the nativist faces a dilemma. Answering 'No' would imply that the concept VERTEBRATE could not be acquired through ostension. For if the concept-forming mechanism were not sensitive to being a vertebrate, then the process by which VERTEBRATE is acquired, and subsequently tokened, could not be subsumed under a primary law relating being a vertebrate to being a cause of VERTEBRATE tokens. VERTEBRATE-acquisition would have to be subsumed under one or other of the secondary laws involving VERTEBRATE, which is to say that VERTEBRATE could only be acquired through reference-borrowing. Since this would seem to be false, Fodor had better say 'Yes', admitting that the property being a vertebrate, as well as being an echidna, can trigger the acquisition mechanism. But if both echidna-hood and vertebrate-hood can trigger the acquisition mechanism, what determines which of these is the causally efficacious property in a given instance? Or, to put it in more Fodorean terms, what determines which law subsumes the acquisition process in case we are experientially connected with an echidna? Is it the primary law relating being an echidna to being a cause of ECHIDNA tokens (in which case the mechanism should output ECHIDNA)? Or is it the primary law relating being a vertebrate with being a cause of VERTEBRATE tokens (in which case the output is VERTEBRATE)? What, in other words, determines which concept gets outputted in response to an input that could cause more than one concept?

### 8.1. Possible responses

In addition to denying outright that being a vertebrate can trigger the acquisition mechanism (a move I argued above to be implausible), there are three further ways Fodor might jump here. First, he might claim that all the properties of the input are causally efficacious, implying that a single echidna-experience occasions acquisition of concepts for all the kinds of which echidnas are members. This is unacceptable: people routinely acquire concepts of particular kinds, without thereby acquiring umpteen others as well. Secondly, he might claim that it is a matter of chance which properties of the input engage the acquisition mechanism: nothing determines whether it's being an echidna or being a vertebrate that is the causally salient property, hence nothing determines which concept is outputted given an echidna-experience. Some concept is acquired; but which concept is entirely up for grabs. This would be true desperation, for it would imply that we should expect to find people who have acquired, say, VERTEBRATE and MAMMAL but not DOG; ARMCHAIR but not CHAIR; MISOGYNIST but not MAN. Typically, though, we first acquire concepts for middle-sized physical objects like CHAIR and DOG, only later acquiring concepts for sub- and superordinate kinds like ARMCHAIR and VERTEBRATE, and later still acquiring abstract concepts like MISOGYNIST. If it truly were a random matter which concepts are acquired given what experiences, in short, we'd expect to find a lot more variability in the course of acquisition than, in fact, we do.

In his (1981), Fodor had a story about why we see the trends we do in concept-acquisition. In explaining why we acquire DOG before acquiring either more general concepts like VERTEBRATE or more particular concepts like POODLE, he suggests that “the triggering structure of the mind is layered, the attainment of some .. concepts being the effect of (hence never prior to) the attainment of others” (1981: 308). Thus, DOG is acquired prior to VERTEBRATE and POODLE because in the latter cases, “triggering .. is mediated by the previously attained conceptual repertoire” (ibid., p. 313); the structure of the acquisition mechanism is such that the acquisition of DOG is a causally necessary condition for the triggering of VERTEBRATE or POODLE, hence VERTEBRATE and POODLE are acquired after DOG. So,

...while all primitive concepts are ipso facto triggered, there is nevertheless a hierarchy of triggers, and it is this hierarchy which predicts the observed order of concept attainment. (1981: 301)

This triggering hierarchy represents a third way the nativist might take out of the difficulties raised in this section. Suppose that the concept-forming mechanism operates according to a rule like this:

Given an echidna-ish input, output ECHIDNA unless ECHIDNA has already been acquired, in which case output MONOTREME; and if MONOTREME has already been acquired, output MAMMAL .. VERTEBRATE .. etc.

Then we could see why experiences of instances of a kind should sometimes be insufficient to occasion concepts of that kind. For, according to this rule, if MAMMAL and MONOTREME have not yet been acquired, VERTEBRATE will not be triggered by an echidna-experience, even though that experience is in fact an experience of a vertebrate.

## 8.2 A dilemma

While representing perhaps the nativist's best shot for solving the present difficulties, postulating a triggering hierarchy also confronts him with something of a dilemma. Suppose that the structure of the hierarchy were relatively simple. Then we would expect to see a lot more invariance in the order of concept-attainment than, arguably, we do. Since mechanisms that vary widely from individual to individual are in general poor candidates for innateness, the claim that the hierarchy is innately specified strongly suggests that it is common to most or all members of our species. Such commonality, indeed, is assumed by Fodor himself when he invokes the hierarchy to account for a relatively stable, species-wide feature of concept attainment, namely, the fact that most children acquire concepts like DOG and CAT before they acquire MAMMAL or POODLE or TABBY.

Unfortunately, however, such species-wide regularities are the exception rather than the rule in acquisition. There's no true generalisation to the effect, for instance, that once a person has PLATYPUS, further platypus-inputs will trigger MONOTREME. There were, after all, plenty of indigenous Australians who saw plenty of platypuses without *ever* acquiring MONOTREME. And quite generally, there seem to be no regularities of this sort apart from the specific case Fodor mentions. This is evidence against a hierarchy, at least of the simple sort we're currently envisaging. For if there were an innately specified triggering hierarchy, we'd expect to see much more invariance in acquisition than there is. Hence, we have the first horn of a dilemma: postulating a simple triggering hierarchy implies interpersonal and cross-cultural conceptual regularities that just do not exist.

In response, Fodor might argue that we don't see such regularities because the hierarchy is much more highly structured than the simplistic rule displayed above suggests.<sup>19</sup> The acquisition of MONOTREME, for example, might require not just the previous acquisition of PLATYPUS, but also of EGG-LAYING, MAMMAL and AUSTRALIAN. Thus platypus-experiences didn't trigger MONOTREME in Australian Aborigines not because there is no hierarchy, but rather because they did not possess the other concepts — like AUSTRALIAN, a much later construction of European imperialism — that the hierarchy specifies as being necessary to MONOTREME's acquisition. The point is quite general. Any empirical facts about who gets which concepts, in what order, in what circumstances, can be accommodated, so long as we're prepared to build enough complexity into the triggering hierarchy.

For that very reason, though, this enrichment strategy also looks suspiciously *ad hoc*: the more bells and whistles we add, the more the hierarchy smells of the 'whatever-it-is-that-explains-concept-acquisition-that's-not-learning', horribly familiar to students of the debate over innate ideas. It also, I fear, looks suspiciously like a step back towards rational-causation, and away from brute-causation. Why do we find it plausible, if we do, to say that prior acquisition of MAMMAL, EGG-LAYING, and

---

<sup>19</sup> David Hilbert and Jim Woodward both suggested this response in comments on an earlier draft.



AUSTRALIAN is necessary for attainment of MONOTREME? I suppose it's because having those concepts is in some intuitive sense necessary for grasping MONOTREME's meaning. But to find such enrichment plausible for this reason is to enjoy intuitions stemming from a rational-causal picture of acquisition, as can be seen by considering that from Fodor's perspective, it might just as well have been XYLOPHONE, CONTRAPUNTAL, and ROMANIAN that turned out to be necessary for getting MONOTREME. Of course the nativist could claim now that it's an empirical question what the structure of the hierarchy is. And of course he'd be right. I nonetheless maintain that postulating a rich innate hierarchy gets whatever plausibility it has from a tacit acceptance of a picture of acquisition that is, from a Fodorean perspective, illicit.

But what's most worrying about this move — what I think leads most directly to the second horn of the nativist's dilemma — is that there is a real tension between viewing the mechanism of acquisition as richly structured, and viewing it as innate. If you're going to be a nativist about the mental, then, if you keep the inbuilt stuff relatively simple, you can get away with issuing a promissory note (or notes) on the biology. You can wave your hands around a bit, claiming either that the mechanism in question was shaped in some relatively straightforward way by natural selection, or else, perhaps, that it arose fortuitously through some non-Darwinian process. The more you enrich the inbuilt structure, though, the more such notes look like so many dud cheques. As nativists about language, among others, have begun to recognize, innate complexity demands an explanation in terms of natural selection.<sup>20</sup> Or, at the very least, it demands more than hand-waving in response to the charge that highly structured innate mechanisms — triggering hierarchies in this case — are nothing more than the somethings-we-know-not-what that explain acquisition.

In short, a simple hierarchy will not account for the facts of acquisition, but a complex hierarchy is evolutionarily implausible. Invoking a triggering hierarchy, I therefore contend, will not illuminate the mechanism of ostensive acquisition. In explaining the nature of the allegedly brute-causal

---

<sup>20</sup> Cf. Oyama (1985); Pinker and Bloom (1990); Pinker (1994), pp. 332ff.

mechanisms mediating the ostensive acquisition of concepts, the Fodor-style nativist is caught between the rock of complete randomness (nothing but chance determines what concept a given experience causes us to acquire) and the hard place of innate over-determination (the triggering hierarchy is the determining factor). Neither of these pictures is in line with what we know to be true about concept-acquisition: there is a certain amount of regularity in the order in which concepts are got, but these regularities occur within a sea of variability. Enriching the hierarchy might explain the observed order-within-disorder of concept-attainment. But, I have argued, this move looks suspiciously like a retreat to a rationalistic picture of acquisition; and, worse, it is biologically implausible.

There is, of course, another way of explaining the facts of concept-acquisition. This is to suppose that acquisition is psychologically mediated, that there is something external to the acquisition mechanism (for example, your background beliefs, desires, and interests) that determines which concept you get from a given experience. If some concepts were acquired 'brutely' in virtue of the way our 'sensorium' is innately structured and others were derived through some collaboration of this 'primitive base' with both our further experience and our surrounding beliefs, needs and interests, then it wouldn't be surprising to observe both some commonalities and a lot of variability in the conceptual history of different individuals. But this last is an account that Fodor cannot accept. To view ostensive acquisition as psychologically mediated is to admit that it is not an innately specified, brute-causal, stimulus-driven process. It is to admit that the radical nativist's acquisition theory is inadequate.

## **9. The mechanism of reference-borrowing**

I argued in Section 8 that, given his strictures as to the brute-causality of acquisition, the Fodorean nativist cannot plausibly account for the ostensive acquisition of concepts. In so arguing, I took Sterelny's *Qua* problem, which he expressed as concerning the determination of reference, and reformulated it as a problem about the order of concept-attainment. Now I take up Sterelny's and

Loar's concerns with reference-borrowing, with the aim again of presenting them in a way that focuses on the psychological mechanisms responsible for acquisition.

I acquired PLATYPUS, let's suppose, through hearing someone utter the word 'platypus'. You acquired it, let's say, through watching a nature programme on TV. An Italian acquires it through reading 'ornitorinco'. A speaker of ASL acquires it through seeing a certain sequence of hand-movements. Assuming that each of us possesses roughly the same kind of genetically-specified acquisition device — assuming, that is, that each of us could, in a different environment, have got PLATYPUS in some way different from that in which we did — we must conceive of a mechanism such that any of these types of experiential input suffices to cause the delivery of the concept PLATYPUS. The mechanism, in other words, must be built such that any of the properties: being a token of the sound /platypus/, being a token TV show about platypuses, being a token of the written word 'ornitorinco', being a hand- movement of so-and-so kind, etc., is enough to brute-cause the output of a mental symbol occupying the place in the nexus of nomological and causal relations constituting the world that (on the asymmetric-dependence theory) determines that concept's reference to platypuses.

But how could a brute-causal mechanism be sensitive to such a diverse range of inputs? It is (just) possible to imagine a machine that is brutally sensitive to being a platypus. But it is very much harder to conceive a brute-causal mechanism that is sensitive to the huge class of properties that, on the view we are now considering, must be triggers for PLATYPUS. For the properties in question are an extremely motley bunch: the written symbol 'ornitorinco' shares no obvious intrinsic properties with the spoken sound /platypus/, which shares no properties with a picture of a platypus, which shares no properties with the ASL sign meaning platypus. What kind of brute-causal mechanism could it be that managed to recognise and respond to just these things? And how, as Sterelny (1989) asks, could such a mechanism have evolved? Recall that on the present view, it's just because of Nature's whim that we are wired up to acquire PLATYPUS when stimulated by platypuses. But how could evolution have 'known' that, in the eighteenth century, a certain sequence of sounds would be introduced to

refer in English to platypuses, or that a certain sequence of squiggles on a page would in Italian designate those animals, or that, in the twentieth century, a very particular series of hand-movements would be introduced in ASL to mean platypus— and thus know to make our minds such that hearing this sound or seeing those squiggles, or witnessing those movements would output the correct concept?

### 9.1 Possible responses

There are again three ways that a Fodorean might respond to this kind of puzzlement. First, he might emphasize that the fact that we can't imagine such a mechanism doesn't mean that there isn't one. And nor does the fact that we can't imagine how it might have been built entail that Mother Nature didn't build it. Indeed, he might continue, his argument shows that there must be and she must have; and all our failure of imagination should do is to remind us that there are more things about the place than we (if not Fodor) ever dreamed of. To make this sort of response, however, is tantamount to admitting that Fodorean nativism is just a redescription of the problem about how we acquire concepts, not a solution to it. The nativist's genetically-determined, brute-causal route from stimuli to concepts degenerates into the 'something we know not what' — mentioned above and satirized by Oyama (1985) as being analogous to an appeal to vital spirits in explaining life — that explains acquisition.

Secondly, he might agree that PLATYPUS-triggers do not share any obvious intrinsic properties and maintain instead that the mechanism is sensitive to some other, less obvious property that those things do share. However it seems that the only property that these triggers share is the relational property representing platypuses. But as Fodor argues in his (1987), pp. 33-41), plausibly enough I think, relational properties of things can only affect their causal powers if there is some natural law

or other mechanism that explains how this effect occurs.<sup>21</sup> If we look for a mechanism by which to relate representing platypuses to PLATYPUS, however, we are back where we started: a mechanism would ground the causal powers of representing platypuses by responding to some intrinsic property or properties shared by things possessing that relational property. But, as we've just seen, no such intrinsic property is available in this case. Perhaps, then, it's a matter of natural law that representing platypuses causes PLATYPUS? But to claim solely on the basis of the present difficulty concerning non-ostensive acquisition that there is a special law of nature in virtue of which representing platypuses causes tokenings of PLATYPUS is outrageously *ad hoc*. In any case, it does not really solve the problem at hand. For what we want to know is *how there could be* a brute-causal mechanism that outputs PLATYPUS given any of the diverse experiences that *de facto* lead to the non-ostensive acquisition of that concept. We have seen that there are no intrinsic properties shared by the inputs that the machine could respond to. So then we asked how a machine could respond 'brutely' to the one (relational) property that the inputs do share. And to be told just that there is a special law of nature in virtue of which that mechanism can respond to that property is, its flagrant *ad hockery* aside, entirely beside the point. What we want to know, when we are talking about acquisition mechanisms, is *how the laws of nature are implemented*. Merely insisting on there being a law, then, is to skirt the problem altogether.

A third way in which the nativist might try to explain the mechanism for non-ostensive acquisition is by again invoking his 'hierarchy of triggers'. As described above (Section 8.1), the hierarchy explained why experiences of instances of concepts are insufficient for acquisition to occur. What needs explaining, though, in case a concept is acquired through reference-borrowing, is not why experiences of instances are sometimes *insufficient* for acquisition, but why they are sometimes *unnecessary*. The question is 'how can you get MONOTREME from a non-monotreme trigger?' and

---

<sup>21</sup> For example, the relational property being a Sagittarius has no causal powers because there is no law or causal mechanism explaining how the positions of the planets at one's birth affect the subsequent course of one's life.

not (as before) 'how can you fail to get MONOTREME despite (like the Aborigines) experiencing lots of monotremes?'

A possible elaboration of the hierarchy is suggested by Fodor's discussion of what he calls the 'mental chemistry' view of concept-acquisition.<sup>22</sup> On this view, according to Fodor, concepts can be 'triggered' by other concepts alone:

...if entertaining A and B is causally sufficient for acquiring C, then ... (a)ll that's required to learn C is the previous acquisition of A and B. That's to say, in effect, that some simple concepts may be triggered by other concepts. (1981: 303-4)

While it is somewhat unclear whether Fodor meant to endorse this view, it could nonetheless provide the basis for extending yet again the triggering hierarchy. For it implies not merely that possession of one concept might be necessary for the acquisition of another; but also that possession of one concept may be sufficient for the attainment of another. If so, then when we acquire PLATYPUS from hearing someone use the word 'platypus', something like the following might occur. Hearing the sound /platypus/ causes ostensive acquisition of the concept /PLATYPUS/, or perhaps WORD MEANING PLATYPUS, or perhaps just PLATYPUS REPRESENTATION; which in turn, and with no further experiential input, brute-causes PLATYPUS. Non-ostensive acquisition, then, would be a two-stage process. Our experience of a representation of an F triggers our concept REPRESENTATION OF AN F, which triggers our concept F-NESS.

But, first, it's unclear how one acquires the concept REPRESENTATION OF A PLATYPUS prior to possessing a concept of the thing, the platypus, that the representation is a representation of. If having PLATYPUS is a necessary condition for having PLATYPUS-REPRESENTATION, then acquisition of PLATYPUS-REPRESENTATION cannot cause acquisition of PLATYPUS. Hence the non-

---

<sup>22</sup> See Fodor (1981: 299ff. Briefly, the idea of mental chemistry seems to be that concepts may be composed out of other concepts without containing them as parts.

ostensive acquisition of PLATYPUS cannot in point of logic be explained by extending the notion of a triggering hierarchy in exactly the way we are considering.

Secondly, and here I return to the issue raised in Section 8.2, it's surely just bizarre to suppose that we could have evolved such that hearing someone talk about platypuses, whether in words or in gestures, could brute-cause any concepts. And it doesn't mitigate the implausibility of this supposition one whit to protest that /platypus/ was recruited to trigger PLATYPUS-REPRESENTATION, not PLATYPUS itself. So if it you find it questionable on evolutionary grounds to attempt to skirt the *Qua* Problem by holding that our minds are structured such that attainment of MAMMAL and EGG-LAYING (let alone XYLOPHONE and CONTRAPUNTAL) are necessary to the acquisition of MONOTREME, you're going to find it even more implausible to think that we evolved such that experiencing one of the myriad representations of platypuses there are in the world suffices for acquiring PLATYPUS-REPRESENTATION; and you're going to find it yet more implausible again-not to say impossible-to suppose that acquisition of PLATYPUS-REPRESENTATION is sufficient for acquisition of PLATYPUS itself.

Which leaves us with a final possibility for dealing with non-ostensive acquisition: denial. The nativist might choose simply to deny that concepts can be acquired through reference-borrowing and insist that the only route to acquisition is ostension. But this, as I've already claimed, is untenable. If an acquisition theory entails that most of us don't have most of our concepts, then so much the worse for it.

And so much the worse for it is, I think, the best we can say about the Fodorean nativist's 'theory' of concept-acquisition. In Section 8, I argued that it fails to explain ostensive acquisition, for a 'brute-causal' mechanism cannot 'know' which concept it should output, given an input that could in principle function to elicit many different concepts. Here, I've been exploring ways that the nativist might attempt to avoid the additional difficulties raised by non-ostensive acquisition. I've suggested two possibilities: that reference- borrowing is a matter of the triggering of concepts by

representations of their instances; or that it is, in virtue of the hierarchy of triggers, mediated by the acquisition of other concepts which themselves serve to make available the borrowed concept. I have argued that neither of these proposals works. That humanly constructed signs could trigger PLATYPUS is in conflict with the nativist's contention that the triggering relation is a brute-causal one engineered by natural selection. And the idea that a triggering hierarchy is implicated is similarly a non-starter. Since acquiring PLATYPUS REPRESENTATION presupposes that we have PLATYPUS available, the latter could not referee the acquisition of the former.

So were philosophers right after all to dismiss Fodor's radical concept-nativism? If our communal neglect of this view was simply a kind of gamble — a bet that something would turn out to be wrong with the theory — then, as the arguments of this paper show, we lucked out. Fodorean nativism can explain neither the ostensive nor the non-ostensive acquisition of concepts— which is to say that it can explain the acquisition of no concepts at all. But if our community's insipid reaction to Fodor's position is evidence instead of a complacent belief in the basic soundness of the foundations of modern psychology, then our reaction has been badly out of line. For finding reasons to reject Fodorean nativism is one thing; dealing with the arguments that provoked it is another. Fodor's (1975) and (1981) arguments look valid enough, and their premises — that learning is a computational process and that learning a concept is a matter of learning what it means— seem perhaps even more uncontroversial today than they did in 1975. So while the rout of mad dog nativism may vindicate our philosophical intuitions, it serves also to underscore that we have on hand neither a definitive account of where Fodor's argument goes wrong, nor a plausible alternative to traditional acquisition theories.<sup>23</sup>

### **Acknowledgements**

I am grateful for the help of David Hilbert, Kim Sterelny, and Jim Woodward in writing this paper.

---

<sup>23</sup> In my (forthcoming), Chs. 4-6, these issues are discussed further.



## References

- Cowie, F. (forthcoming): What's Within? Nativism Reconsidered, New York, Oxford University Press.
- Devitt, M. (1981): Designation, New York, Columbia University Press.
- Devitt, M. and Sterelny, K. (1987): Language and Reality, Cambridge, MA, MIT Press/ Bradford Books.
- Dretske, F. (1981): Knowledge and the Flow of Information, Cambridge, MA, MIT Press.
- Fodor, J. A. (1975): The Language of Thought, New York, Crowell.
- Fodor, J. A. (1979): 'On the Impossibility of Acquiring "More Powerful" Structures', in M. Piattelli-Palmarini (ed.), Language and Learning: The Debate between Jean Piaget and Noam Chomsky, London, Routledge & Kegan Paul, pp. 142-62.
- Fodor, J. A. (1981): 'The Present Status of the Innateness Controversy', in J. A. Fodor, RePresentations: Philosophical Essays on the Foundations of Cognitive Science, Cambridge, MA, MIT Press/Bradford Books, pp. 257-316.
- Fodor, J. A. (1983): The Modularity of Mind, Cambridge, MA, MIT Press/Bradford Books.
- Fodor, J. A. (1984a): 'Observation Reconsidered', Philosophy of Science, 51, pp. 23-43, reprinted in Fodor (1990a), pp. 231-51.
- Fodor, J. A. (1984b): 'Semantics, Wisconsin Style', Synthese, 59, pp. 231-50, reprinted in Fodor (1990a), pp. 31-50.
- Fodor, J. A. (1987): Psychosemantics: The Problem of Meaning in the Philosophy of Mind, Cambridge, MA, MIT Press/Bradford Books.
- Fodor, J. A. (1990a): A Theory of Content and Other Essays, Cambridge, MA, MIT Press/Bradford Books.

- Fodor, J. A. (1990b): 'Psychosemantics: Or Where Do Truth Conditions Come From?' in W. G. Lycan (ed.), *Mind and Cognition: A Reader*, Oxford, Blackwell.
- Fodor, J. A. (1991): 'Replies', in B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and His Critics*, Oxford, UK, and Cambridge, MA, Blackwell.
- Fodor, J. A. (1998): *Concepts: Where Cognitive Science Went Wrong*, Oxford, Oxford University Press.
- Fodor, J. A. and Pylyshyn, Z. W. (1988): 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition*, 28, pp. 3-71.
- Kripke, S. (1972/1980): *Naming and Necessity*, Cambridge, MA, Harvard University Press.
- Loar, B. (1991): 'Can We Explain Intentionality?' in B. Loewer and G. Rey (eds), *Meaning in Mind: Fodor and his Critics*, Oxford, UK, and Cambridge, MA, Blackwells.
- Oyama, S. (1985): *The Ontogeny of Information*, Cambridge, Cambridge University Press.
- Pinker, S. (1994): *The Language Instinct: How the Mind Creates Language*, New York, Harper.
- Pinker, S. and Bloom, P. (1990): 'Natural Language and Natural Selection', *Behavioral and Brain Sciences*, 13, pp. 707-84.
- Putnam, H. (1975): 'The Meaning of "Meaning" ', in *Mind, Language and Reality: Philosophical Papers*, Vol. 2, Cambridge, Cambridge University Press, pp. 33-69.
- Smolensky, P. (1988): 'On the Proper Treatment of Connectionism', *Behavioral and Brain Sciences*, 11, pp. 1-74.
- Sterelny, K. (1983): 'Natural Kind Terms', *Pacific Philosophical Quarterly*, 64, pp. 110-25.
- Sterelny, K. (1989): 'Fodor's Nativism', *Philosophical Studies*, 55, pp. 119-41.